Mathematical Programming manuscript No.
(will be inserted by the editor)

Igor Griva · Roman A. Polyak

Primal-dual nonlinear rescaling method with dynamic scaling parameter update

Received: / Revised version:

Abstract. In this paper we developed a general primal-dual nonlinear rescaling method with dynamic scaling parameter update (PDNRD) for convex optimization. We proved the global convergence, established 1.5-Q-superlinear rate of convergence under the standard second order optimality conditions. The PDNRD was numerically implemented and tested on a number of nonlinear problems from COPS and CUTE sets. We present numerical results, which strongly corroborate the theory.

Key words. Nonlinear rescaling, duality, Lagrangian, primal-dual, multipliers method.

1. Introduction

The success of the primal-dual (PD) methods for linear programming (see [16]-[18], [32]-[34] and references therein) has stimulated substantial interest in the primal-dual approach for nonlinear programming calculations (see [10], [11], [22], [30]). The best known approach for developing primal-dual methods is based on path-following ideology (see e.g. [30]). It requires an unbounded increase of the scaling parameter to guarantee the convergence. Another approach (see [29]) is based on the nonlinear rescaling (NR) methods (see [24]-[28]). The NR method does not require an unbounded increase of the scaling parameter because it has an extra tool to control convergence: the Lagrange multipliers vector. Each step of the NR method alternates the unconstrained minimization of the Lagrangian for the equivalent problem with the Lagrange multipliers update. The scaling parameter can be fixed or updated from step to step. Convergence of the NR method under the fixed scaling parameter allows avoiding the ill-conditioning of

Igor Griva: Department of SEOR, George Mason University, Fairfax VA 22030

Roman A. Polyak: Department of SEOR and Mathematical Sciences Department, George Mason University, Fairfax VA 22030

the Hessian of the minimized function. Moreover, under the standard second order optimality conditions the NR methods converge with Q-linear rate for any fixed but large enough scaling parameter.

To improve the rate of convergence one has to increase the scaling parameter from step to step. Again, it leads to the ill-conditioned Hessian of the minimized function and to a substantial increase of the computational work per Lagrange multipliers update. Therefore, in [29] the authors introduced and analyzed the primal-dual NR method (PDNR). The unconstrained minimization and the Lagrange multipliers update are replaced with solving the primal-dual system using Newton's method. Under the standard second order optimality conditions, the PDNR method converges with linear rate. Moreover, for any given factor $0 < \gamma < 1$ there exists such a fixed scaling parameter that from some point on just one Newton step shrinks the distance between the current approximation and the primal-dual solution by factor γ.

In this paper we show that the rate of convergence can be substantially improved without compromising computational effort per step. The improvement is achieved by increasing in a special way the scaling parameter from step to step.

The fundamental difference between the PDNR and the Newton NR methods ([19], [26]) lies in the fact, that in the final phase of the computational process the former does not perform the unconstrained minimization at each step. Therefore the ill-conditioning becomes irrelevant for the PDNR method while for the Newton NR method it leads to numerical difficulties. Moreover, the drastic increase of the scaling parameter in the final stage makes the primal-dual Newton direction close to the corresponding direction obtained by Newton's method for solving the Lagrange system of equations that corresponds to the active constraints. This is critical for improving the rate of convergence.

Our first contribution is the globally convergent primal-dual nonlinear rescaling method with dynamic scaling parameter update (PDNRD). Our second contribution is the proof that the PDNRD method with a special scaling parameter update converges with 1.5-Q-superlinear rate under the standard second order optimality conditions. Our third contribution is the MATLAB based code, which has been tested on a large number of NLP including problems from COPS [5] and CUTE [6] sets. The obtained numerical results corroborate the theory and show that the PD approach in the NR framework has a good potential to become a competitive tool in the NLP area.

The paper is organized as follows. In the next section we consider the convex optimization problem with inequality constraints and discuss the basic assumptions. In section 3 we describe the classical NR multipliers method which has the Q-linear rate of convergence under the fixed scaling parameter. In section 4 we describe and study the PDNRD method. Our main focus is the asymptotic 1.5-Q-superlinear rate of convergence of the PDNRD method. Section 5 describes the globally convergent PDNRD method together with its numerical realization. Section 6 shows the numerical results obtained by testing the PDNRD method. We conclude the paper by discussing issues related to future research.

2. Statement of the problem and basic assumptions

Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be convex, all $c_i : \mathbb{R}^n \to \mathbb{R}^1$, $i = 1, \ldots, m$ be concave and smooth functions. We consider a convex set $\Omega = \{x \in \mathbb{R}^n : c_i(x) \geq 0,\ i = 1, \ldots, m\}$ and the following convex optimization problem

$$(\mathcal{P}) \qquad x^* \in X^* = Argmin\{f(x) | x \in \Omega\}.$$

We assume that:

A. The optimal set X^* is not empty and bounded.

B. The Slater's condition holds, i.e. there exists $\hat{x} \in \mathbb{R}^n : c_i(\hat{x}) > 0$, $i = 1, \ldots, m$.

Due to the assumption **B**, the Karush-Kuhn-Tucker's (K-K-T's) conditions hold true, i.e. there exists a vector $\lambda^* = (\lambda_1^*, ..., \lambda_m^*) \in \mathbb{R}^m_+$ such that for the Lagrangian $L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i c_i(x)$ we have

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, \tag{2.1}$$

and the complementary slackness conditions

$$\lambda_i^* c_i(x^*) = 0,\ i = 1, \ldots, m \tag{2.2}$$

hold true.

Let us assume that the active constraint set at x^* is $I^* = \{i : c_i(x^*) = 0\} = \{1, \ldots, r\}$. We consider the vectors functions $c^T(x) = (c_1(x), \ldots, c_m(x))$, $c^T_{(r)}(x) = (c_1(x), \ldots, c_r(x))$, and their Jacobians $\nabla c(x) = J(c(x))$ and $\nabla c_{(r)}(x) = J(c_{(r)}(x))$. The sufficient regularity conditions

$$rank \nabla c_{(r)}(x^*) = r, \lambda_i^* > 0,\ i \in I^* \tag{2.3}$$

together with the sufficient condition for the minimum x^* to be isolated

$$(\nabla^2_{xx} L(x^*, \lambda^*)y, y) \geq \rho(y, y),\ \rho > 0, \forall y \neq 0 : \nabla c_{(r)}(x^*)y = 0 \tag{2.4}$$

comprise the standard second order optimality conditions.

3. Equivalent problem and nonlinear rescaling method

Let $-\infty < t_0 < 0 < t_1 < \infty$. We consider a class Ψ of twice continuously differential functions $\psi : (t_0, t_1) \to \mathbb{R}$, which satisfy the following properties

1^0. $\psi(0) = 0$.

2^0. $\psi'(t) > 0$.

3^0. $\psi'(0) = 1$.

4^0. $\psi''(t) < 0$.

5^0. there is $a > 0$ that $\psi(t) \leq -at^2$, $t \leq 0$.

6^0. a) $\psi'(t) \leq b_1 t^{-1}$, b) $-\psi''(t) \leq b_2 t^{-2}$, $t > 0$, $b_1 > 0$, $b_2 > 0$.

Let us consider a few transformations $\psi \in \Psi$.

1. Exponential transformation [15]

$$\psi_1(t) = 1 - e^{-t}.$$

2. Logarithmic MBF [24]

$$\psi_2(t) = \ln(t + 1).$$

3. Hyperbolic MBF [24]

$$\psi_3(t) = \frac{t}{1 + t}.$$

Each of the above transformations can be modified in the following way. For a given $-1 < \tau < 0$ we define quadratic extrapolation of the transformations 1. − 3. by formulas

4.

$$\psi_{q_i}(t) = \begin{cases} \psi_i(t), & t \geq \tau, \\ q_i(t) = a_i t^2 + b_i t + c_i, & t \leq \tau. \end{cases}$$

where $a_i,\ b_i,\ c_i$ we find from the following equations: $\psi_i(\tau) = q_i(\tau),\ \psi_i'(\tau) = q_i'(\tau),\ \psi_i''(\tau) = q_i''(\tau)$. We obtain $a = 0.5\psi''(\tau)$, $b = \psi'(\tau) - \tau\psi''(\tau)$, $c = \psi(\tau) - \tau\psi'(\tau) + \tau^2\psi''(\tau)$, so $\psi_{q_i}(t) \in C^2$. Such modification of logarithmic MBF was introduced in [3] and successfully used for solving large-scale NLP (see [2], [3], [7], [20]). It is easy to check that transformations 1. − 4. satisfy properties $1^0 - 6^0$.

Modification 4. leads to transformations, which are defined on $(-\infty, \infty)$ and along with penalty function properties, have some extra important features. One can find other examples of transformations with similar properties in [1], [3], [25].

For any given transformation $\psi \in \Psi$ and any $k > 0$ due to $1^0 - 3^0$ we obtain

$$\Omega = \left\{x : k^{-1}\psi(kc_i(x)) \geq 0, i = 1, \ldots, m\right\}. \tag{3.1}$$

Therefore for any $k > 0$ the following problem

$$x^* \in X^* = Argmin\{f(x) | k^{-1}\psi(kc_i(x)) \geq 0,\ i = 1, \ldots, m\} \tag{3.2}$$

is equivalent to the original problem $\mathcal{P}$. The Classical Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^1_{++} \to \mathbb{R}^1$ for the equivalent problem (3.2)

$$\mathcal{L}(x, \lambda, k) = f(x) - k^{-1} \sum_{i=1}^{m} \lambda_i \psi(kc_i(x)). \tag{3.3}$$

is our main tool.

Let $\lambda^0 \in \mathbb{R}^m_{++}$ be the initial Lagrange multipliers vector. We consider a monotone increasing sequence $\{k_s\} : k_0 > 0, \lim_{s\to\infty} k_s = \infty$ of scaling parameters. The NR method with the dynamic scaling parameter update for a given scaling parameters sequence $\{k_s\}$ generates the primal-dual sequence $\{x^s, \lambda^s\}$ as follows. Let us assume that a primal-dual pair $(x^s, \lambda^s) \in \mathbb{R}^n \times \mathbb{R}^m_{++}$ have been already found, we find the next approximation (x^{s+1}, λ^{s+1}) by the following formulas

$$x^{s+1} = argmin\left\{\mathcal{L}(x, \lambda^s, k_s) \,|\, x \in \mathbb{R}^n\right\}, \tag{3.4}$$

or

$$x^{s+1}:\nabla_x\mathcal{L}(x^{s+1},\lambda^s,k_s)=\nabla f(x^{s+1})-\sum_{i=1}^{m}\psi'\left(k_sc_i(x^{s+1})\right)\lambda_i^s\nabla c_i(x^{s+1})=0, \quad (3.5)$$

and

$$\lambda_i^{s+1}=\psi'\left(k_sc_i(x^{s+1})\right)\lambda_i^s, i=1,\ldots,m. \quad (3.6)$$

or

$$\lambda^{s+1}=\Psi'\left(k_sc(x^{s+1})\right)\lambda^s, \quad (3.7)$$

where $\Psi'\left(k_sc(x^{s+1})\right)=diag\left(\psi'\left(k_sc_i(x^{s+1})\right)\right)_{i=1}^m$.

The Nonlinear Rescaling (NR) method (3.4)–(3.7) is well defined due to the properties 4^0-6^0 of transformation $\psi(t)$, convexity of the original problem $\mathcal{P}$ and assumption **A** (see [1]).

Under the standard second order optimality conditions (2.4)–(2.5) the trajectory $\{x^s,\lambda^s\}$ generated by (3.4)–(3.7) converges to the primal-dual solution with superlinear rate when $k_0>0$ is large enough (see ([24], [29]). In other words the following bounds hold

$$\|x^{s+1}-x^*\|\le ck_s^{-1}\|\lambda^s-\lambda^*\|, \quad \|\lambda^{s+1}-\lambda^*\|\le ck_s^{-1}\|\lambda^s-\lambda^*\|, \quad (3.8)$$

where $c>0$ is independent from $\{k_s\}$.

Finding x^{s+1} requires solving an unconstrained minimization problem (3.4), which is generally speaking an infinite procedure. The stopping criteria (see [26], [29]) allows to replace x^{s+1} by an approximation $\bar{x}^{s+1}$, which can be found in a finite number of Newton steps. If $\bar{x}^{s+1}$ is used in the formula (3.7) for the Lagrange multipliers update then bounds similar to (3.8) remain true.

Let us consider the sequence $\{\bar{x}^s,\bar{\lambda}^s\}$ generated by the following formulas

$$\bar{x}^{s+1}\ :\ \|\nabla_x\mathcal{L}(\bar{x}^{s+1},\bar{\lambda}^s,k_s)\|\le\sigma k_s^{-1}\|\Psi'\left(k_sc(\bar{x}^{s+1})\right)\bar{\lambda}^s-\bar{\lambda}^s\|, \quad (3.9)$$

$$\bar{\lambda}^{s+1}=\Psi'\left(k_sc(\bar{x}^{s+1})\right)\bar{\lambda}^s \quad (3.10)$$

for some initial vector $\lambda^0\in\mathbb{R}_{++}^m$ of Lagrange multipliers and positive monotone increasing sequence $\{k_s\}$ with $k_0>0$ large enough.

By using considerations similar to those in [27] we can prove the following proposition.

Proposition 1. *If the standard second order optimality conditions hold and the Hessians $\nabla^2 f(x)$ and $\nabla^2 c_i(x)$, $i = 1, \ldots, m$ satisfy the Lipschitz conditions*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_0\|x - y\|, \; \|\nabla^2 c_i(x) - \nabla^2 c_i(y)\| \leq L_i\|x - y\|, \; i = 1, \ldots, m, \tag{3.11}$$

then there is $k_0 > 0$ large enough, that for the primal-dual sequence $\{\bar{x}^s, \bar{\lambda}^s\}$ generated by formulas (3.9)-(3.10) the following estimations hold true and $c > 0$ is independent from k_s for all $k_s \geq k_0$.

$$\|\bar{x}^{s+1} - x^*\| \leq c(1+\sigma)k_s^{-1}\|\bar{\lambda}^s - \lambda^*\|, \quad \|\bar{\lambda}^{s+1} - \lambda^*\| \leq c(1+\sigma)k_s^{-1}\|\bar{\lambda}^s - \lambda^*\|. \tag{3.12}$$

To find an approximation $\bar{x}^{s+1}$ we use Newton's method for solving the primal-dual system that we consider in the next section. Generally speaking it requires several Newton steps to find $\bar{x}^{s+1}$. Then we update the Lagrange multipliers using $\bar{x}^{s+1}$ instead of x^{s+1} in (3.7). The PDNRD method follows NR trajectory $(\bar{x}^s, \bar{\lambda}^s)$ until the primal-dual approximation reaches the neighborhood of the primal-dual solution. Then the PDNRD method turns into Newton's method for the Lagrange system of equations corresponding to the active constraints. From this point on it requires only one Newton step to obtain 1.5-Q-superlinear rate of convergence.

4. Primal-dual NR method with dynamic scaling parameter update

In this section we describe and analyze the PDNR method with dynamic scaling parameter update (PDNRD).

In the following we use the vector norm $\|x\| = \max_{1\leq i\leq n} |x_i|$, the matrix norm $\|Q\| = \max\limits_{1\leq i\leq n} \left(\sum\limits_{j=1}^{n} |q_{ij}| \right)$.

To measure a distance between the current approximation (x, λ) and the solution we introduce a merit function:

$$\nu(x, \lambda) = \max \left\{ \|\nabla_x L(x, \lambda)\|, \; -\min_{1\leq i\leq m} c_i(x), \; \sum_{i=1}^{m} |\lambda_i||c_i(x)|, \; -\min_{1\leq i\leq m} \lambda_i, \right\}. \tag{4.1}$$

For a given $x \in \mathbb{R}^n$, Lagrange multipliers vector $\lambda \in \mathbb{R}^m_+$ and the scaling parameter $k > 0$ one step of the NR method is equivalent to solving the following primal-dual system

$$\nabla_x \mathcal{L}(\hat{x}, \lambda, k) = \nabla f(\hat{x}) - \sum_{i=1}^{m} \psi' \left(kc_i(\hat{x})\right) \lambda_i \nabla c_i(\hat{x}) = \nabla_x L(\hat{x}, \hat{\lambda}) = 0, \tag{4.2}$$

$$\hat{\lambda} = \Psi'\left(kc(\hat{x})\right)\lambda, \tag{4.3}$$

for $\hat{x}$ and $\hat{\lambda}$, where $\Psi'(kc(\hat{x})) = diag(\psi(kc_i(\hat{x})))_{i=1}^m$.

We are going to update the scaling parameter $k > 0$ at each step according to the following formula

$$\hat{k} = \nu(\hat{x}, \hat{\lambda})^{-0.5}. \tag{4.4}$$

Let us first consider the Primal-Dual system

$$\nabla_x L(\hat{x}, \hat{\lambda}) = \nabla f(\hat{x}) - \sum_{i=1}^m \hat{\lambda}_i \nabla c_i(\hat{x}) = 0, \tag{4.5}$$

$$\hat{\lambda} = \Psi'\left(kc(\hat{x})\right)\lambda. \tag{4.6}$$

We apply Newton's method for solving system (4.5)–(4.6) for $\hat{x}$ and $\hat{\lambda}$ using (x, λ) as a starting point.

Assuming that $\hat{x} = x + \Delta x$, $\hat{\lambda} = \lambda + \Delta\lambda$, and by linearizing (4.5) - (4.6) we obtain the following system for finding the Primal-Dual Newton direction $(\Delta x, \Delta\lambda)$

$$\nabla f(x) + \nabla^2 f(x)\Delta x - \sum_{i=1}^m (\lambda_i + \Delta\lambda_i)(\nabla c_i(x) + \nabla^2 c_i(x)\Delta x) = 0, \tag{4.7}$$

$$\lambda_i + \Delta\lambda_i = \psi'\left(k(c_i(x) + \nabla c_i^T(x)\Delta x)\right)\lambda_i =$$

$$\psi'\left(kc_i(x)\right)\lambda_i + k\psi''\left(kc_i(x)\right)\lambda_i \nabla c_i^T(x)\Delta x,\ i = 1, \dots, m. \tag{4.8}$$

Ignoring terms of the second and higher orders and assuming that $\bar{\lambda} = \Psi'\left(kc(x)\right)\lambda$ we can rewrite (4.7)–(4.8) as follows

$$\nabla^2_{xx} L(x, \lambda)\Delta x - \nabla c(x)^T \Delta\lambda = -\nabla_x L(x, \lambda) = -\nabla_x L(\cdot),$$

$$-k\Lambda\Psi''\left(kc(x)\right)\nabla c(x)\Delta x + \Delta\lambda = \bar{\lambda} - \lambda,$$

or

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T(\cdot) \\ -k\Lambda\Psi''\left(\cdot\right)\nabla c(\cdot) & I_m \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \bar{\lambda} - \lambda \end{bmatrix}, \tag{4.9}$$

where $\nabla c(\cdot) = \nabla c(x)$, $\Psi''(\cdot) = \Psi''\left(kc(x)\right) = diag\ \left(\psi''\left(kc_i(x)\right)\right)_{i=1}^m$, $\Lambda = diag\ (\lambda_i)_{i=1}^m$ and I_m is an identity matrix in $\mathbb{R}^{m,m}$. By introducing

$$N(\cdot) = \begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T(\cdot) \\ -k\Lambda\Psi''(\cdot)\nabla c(\cdot) & I_m \end{bmatrix},$$

we can rewrite system (4.9) as

$$N(\cdot)\begin{bmatrix}\Delta x\\ \Delta\lambda\end{bmatrix}=\begin{bmatrix}-\nabla_x L(\cdot)\\ \bar\lambda-\lambda\end{bmatrix}.$$

To make the matrix $N(\cdot)$ nonsingular for any (x,λ) we regularize the Hessian of the Lagrangian

$$N_k(\cdot)=\begin{bmatrix}\nabla^2_{xx}L(\cdot)+\frac{1}{k^2}I_n & -\nabla c^T(\cdot)\\ -k\Lambda\Psi''(\cdot)\nabla c(\cdot) & I_m\end{bmatrix}, \tag{4.10}$$

where I_n is an identity matrix in $\mathbb{R}^{n,n}$. The reason for choosing such a regularization parameter will become clear later from the convergence proof. The choice guarantees global convergence and does not compromise the rate of convergence.

For a given $x\in\mathbb{R}^n$, Lagrange multipliers vector $\lambda\in\mathbb{R}^m_+$ and scaling parameter $k>0$ one step of the PDNRD method consists of the following operations:

1. Find the primal-dual Newton direction from the system

$$N_k(\cdot)\begin{bmatrix}\Delta x\\ \Delta\lambda\end{bmatrix}=\begin{bmatrix}-\nabla_x L(\cdot)\\ \bar\lambda-\lambda\end{bmatrix}. \tag{4.11}$$

2. Find the new primal-dual vector

$$\hat x:=x+\Delta x,\quad \hat\lambda:=\lambda+\Delta\lambda. \tag{4.12}$$

3. Update the scaling parameters

$$\hat k=\nu(\hat x,\hat\lambda)^{-0.5}. \tag{4.13}$$

The matrix $N_k(\cdot)$ is often sparse, therefore the sparse numerical linear algebra technique can be very efficient (see i.e. [21], [31]).

The following lemma guarantees that the method (4.11)–(4.13) is well defined.

Lemma 1. *Matrix $N_k(x,\lambda)$ is nonsingular for any primal-dual vector $(x,\lambda)\in\mathbb{R}^n\times\mathbb{R}^m_+$ and positive scaling parameter $k>0$.*

Proof. We are going to show that equation $N_k(\cdot)w = 0$, where $w = (u, v)$ implies $w = 0$ for any pair (x, λ) and scaling parameter $k > 0$. We can rewrite the system

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) + \frac{1}{k^2} I_n & -\nabla c^T(\cdot) \\ -k\Lambda\Psi''(\cdot)\,\nabla c(\cdot) & I_m \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

as follows

$$\left(\nabla^2_{xx} L(x, \lambda) + \frac{1}{k^2} I_n\right) u - \nabla c(x)^T v = 0, \tag{4.14}$$

$$-k\Lambda\Psi''(kc(x))\,\nabla c(x)u + v = 0. \tag{4.15}$$

By substituting the value of v from (4.15) into (4.14) we obtain the system

$$\left(\nabla^2_{xx} L(x, \lambda) + \frac{1}{k^2} I_n - k\nabla c^T(x)\Psi''(kc(x))\,\Lambda\nabla c(x)\right) u = 0. \tag{4.16}$$

Due to the convexity of the original problem $\mathcal{P}$ and property 4^0 the matrix

$$M(\cdot) = \nabla^2_{xx} L(x, \lambda) + \frac{1}{k^2} I_n - k\nabla c^T(x)\Psi''(kc(x))\,\Lambda\nabla c(x)$$

is positive definite for any $k > 0$. Therefore from (4.16) we have $u = 0$ and, consequently, due to (4.15) $v = 0$. Lemma 1 is proven.

Let $\Omega_\varepsilon = \{y = (x, \lambda) \,|\, \|y - y^*\| \le \varepsilon\}$. We remind that $I^* = \{1, \ldots, r\}$ and $I^0 = \{r + 1, \ldots, m\}$ are active and inactive sets of constraints respectively. Let us also consider vector-functions $c_{(r)}(x)$, $c_{(m-r)}(x)$, their Jacobians $\nabla c_{(r)}(x)$, $\nabla c_{(m-r)}(x)$ and the Lagrange multipliers vectors $\lambda_{(m-r)}$, $\lambda_{(r)}$, corresponding to the active and inactive sets respectively. Also $L_{(r)}(x, \lambda_{(r)}) = f(x) - \lambda^T_{(r)} c_{(r)}(x)$ is the Lagrangian corresponding to the active set.

The following lemmas take place.

Lemma 2. *Let matrix $A \in \mathbb{R}^{n,n}$ be nonsingular and $\|A^{-1}\| \le M$. Then for any matrix $B \in \mathbb{R}^{n,n}$ such that $\|A - B\| \le \varepsilon$ and $\varepsilon > 0$ small enough matrix B is nonsingular and the following bound holds*

$$\|B^{-1}\| \le 2M.$$

Proof. Since the matrix A is nonsingular, we have

$$B = A - (A - B) = A(I - A^{-1}(A - B)).$$

Let us denote matrix $C = A^{-1}(A - B)$. Since $\|A^{-1}\| \le M$, we can choose such $\varepsilon > 0$ small enough that

$$\|C\|_2 \le \frac{1}{2\sqrt{n}}.$$

Therefore there exists matrix $(I - C)^{-1}$ and we have

$$\|(I - C)^{-1}\| \le \|I\| + \|C\| + \|C\|^2 + \cdots \le 1 + \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^2 + \cdots \le 2.$$

Thus we have the following estimate

$$\|B^{-1}\| = \|(I - C)^{-1}A^{-1}\| \le \|(I - C)^{-1}\|\|A^{-1}\| \le 2M.$$

Lemma is proven.

It follows from (2.3) ans (2.4) (see [23]) that the matrix

$$A = \begin{bmatrix} \nabla^2_{xx} L_{(r)}(x^*, \lambda^*_{(r)}) & -\nabla c^T_{(r)}(x^*) \\ \nabla c_{(r)}(x^*) & 0 \end{bmatrix}$$

has an inverse and there is $M > 0$ such that

$$\|A^{-1}\| \le M. \tag{4.17}$$

Lemma 3. *If the standard second order optimality conditions (2.3)-(2.4) and the Lipschitz conditions (3.11) for Hessians $\nabla^2 f(x)$, $\nabla^2 c_i(x)$, $i = 1, \ldots, m$ are satisfied then there exists such $\varepsilon_0 > 0$ small enough that for any primal-dual pair $y = (x, \lambda) \in \Omega_{\varepsilon_0}$ the following hold true*

1) There exist $0 < L_1 < L_2$ such that the merit function $\nu(y)$ yields

$$L_1\|y - y^*\| \le \nu(y) \le L_2\|y - y^*\|. \tag{4.18}$$

2) Let $D = diag(d_i)_{i=1}^r$ be a diagonal matrix with nonnegative bounded from above elements, i.e. $\max\{d_i\}_{i=1}^r = \bar{d} < \infty$. Then there exists $k_0 > 0$ such that for any $k \ge k_0$ and any $y \in \Omega_{\varepsilon_0}$ the matrices

$$A(x, \lambda_{(r)}) = \begin{bmatrix} \nabla^2_{xx} L_{(r)}(x, \lambda_{(r)}) & -\nabla c^T_{(r)}(x) \\ \nabla c_{(r)}(x) & 0 \end{bmatrix} \quad \text{and} \quad B_k(x, \lambda) = \begin{bmatrix} \nabla^2_{xx} L(x, \lambda) + \frac{1}{k^2} I_n & -\nabla c^T_{(r)}(x) \\ \nabla c_{(r)}(x) & \frac{1}{k} D \end{bmatrix}$$

are nonsingular and the following bound holds

$$\max\left\{\|A^{-1}(x,\lambda_{(r)})\|,\|B_k^{-1}(x,\lambda)\|\right\}\leq 2M. \tag{4.19}$$

Proof. 1) Keeping in mind that $\nu(y^*)=0$ the right inequality (4.18) follows from Lipschitz conditions (3.11) and the boundedness of Ω_{ε_0}. Therefore there exists $L_2>0$ such that

$$\nu(y)\leq L_2\|y-y^*\|.$$

From a definition of a merit function (4.1) we have

$$\|\nabla_x L(x,\lambda)\|\leq \nu(y), \tag{4.20}$$

$$-\min_{1\leq i\leq m} c_i(x)\leq \nu(y), \tag{4.21}$$

$$|\lambda_i||c_i(x)|\leq \nu(y),\quad i=1,\ldots,m. \tag{4.22}$$

Due to the standard second order optimality conditions there exists $\tau_1>0$ such that $c_i(x)\geq\tau_1$, $i\in I^0$, if $y\in\Omega_{\varepsilon_0}$. Therefore from (4.22) we get

$$|\lambda_i|\leq\frac{1}{\tau_1}\nu(y)=C_1\nu(y),\quad i\in I^0, \tag{4.23}$$

where $C_1=\frac{1}{\tau_1}$. Due to the boundedness Ω_{ε_0} there exists also $\tau_2>0$ such that $\|\nabla c_{(m-r)}(x)\|\leq\tau_2$ if $y\in\Omega_{\varepsilon_0}$. Thus taking into account (4.20) we have

$$\|\nabla_x L_{(r)}(x,\lambda_{(r)})\|\leq\|\nabla_x L(x,\lambda)\|+\|\nabla c^T_{(m-r)}(x)\lambda_{(m-r)}\|\leq C_2\nu(y), \tag{4.24}$$

where $C_2=1+(m-r)C_1\tau_2$. Also due to the standard second order optimality conditions there exists $\tau_3>0$ such that $\lambda_i\geq\tau_3$ for $i\in I^*$ if $y\in\Omega_{\varepsilon_0}$. Combining (4.21) and (4.22) we obtain

$$\|c_{(r)}(x)\|\leq C_3\nu(y), \tag{4.25}$$

where $C_3=\min\{1,\frac{1}{\tau_3}\}$.

Let us linearize $\nabla_x L_{(r)}(x,\lambda_{(r)})$ and $c_{(r)}(x)$ at the solution $(x^*,\lambda^*_{(r)})$.

$$\begin{aligned}&\nabla_x L_{(r)}(x,\lambda_{(r)})=\\&\nabla_x L_{(r)}(x^*,\lambda^*_{(r)})+\nabla^2_{xx}L_{(r)}(x^*,\lambda^*_{(r)})(x-x^*)-\nabla c^T_{(r)}(x^*)(\lambda_{(r)}-\lambda^*_{(r)})+\mathcal{O}_{(n)}\|x-x^*\|^2,\end{aligned} \tag{4.26}$$

$$c_{(r)}(x) = c_{(r)}(x^*) + \nabla c_{(r)}(x^*)(x - x^*) + \mathcal{O}_{(r)}\|x - x^*\|^2. \tag{4.27}$$

Keeping in mind K-K-T conditions we can rewrite (4.26)–(4.27) in a matrix form

$$\begin{bmatrix} \nabla^2_{xx} L_{(r)}(x^*, \lambda^*) & -\nabla c^T_{(r)}(x^*) \\ \nabla c_{(r)}(x^*) & 0 \end{bmatrix} \begin{bmatrix} x - x^* \\ \lambda_{(r)} - \lambda^*_{(r)} \end{bmatrix} = \begin{bmatrix} \nabla_x L_{(r)}(x, \lambda_{(r)}) + \mathcal{O}_{(n)}\|x - x^*\|^2 \\ c_{(r)}(x) + \mathcal{O}_{(r)}\|x - x^*\|^2 \end{bmatrix} \tag{4.28}$$

Due to the standard second order optimality conditions the matrix

$$A(x^*, \lambda^*_{(r)}) = \begin{bmatrix} \nabla^2_{xx} L_{(r)}(x^*, \lambda^*_{(r)}) & -\nabla c^T_{(r)}(x^*) \\ \nabla c_{(r)}(x^*) & 0 \end{bmatrix}$$

is nonsingular (see e.g. [23]) and there exists $M > 0$ such that $\|A^{-1}(x^*, \lambda^*_{(r)})\| \le M$. Hence from (4.28) we have

$$\left\| \begin{matrix} x - x^* \\ \lambda_{(r)} - \lambda^*_{(r)} \end{matrix} \right\| \le M \max\{C_2, C_3\}\nu(y) + \mathcal{O}\|y - y^*\|^2.$$

Using (4.23) and assuming $1/L_1 = \max\{C_1, 2M \max\{C_2, C_3\}\}$ we obtain left inequality (4.18), i.e.

$$L_1\|y - y^*\| \le \nu(y).$$

2) The bound (4.19) is a direct consequence of (4.17), Lemma 2 and the Lipschitz conditions (3.11). Lemma 3 is proven.

We are ready to prove the main result. For the method (4.11)–(4.13) the following theorem holds.

Theorem 1. *If the standard second order optimality conditions (2.3)-(2.4) and the Lipschitz conditions (3.11) are satisfied then there exists* $\varepsilon_0 > 0$ *small enough such that for any primal-dual pair* $y = (x, \lambda) \in \Omega_{\varepsilon_0}$ *only one step of PDNRD method (4.11)-(4.13) is required to obtain the new primal-dual approximation* $(\hat{x}, \hat{\lambda})$ *that the following estimation*

$$\|\hat{y} - y^*\| \le C\|y - y^*\|^{\frac{3}{2}} \tag{4.29}$$

holds and $C > 0$ *is a constant depending only on the problem data.*

Proof. Let $\varepsilon_0 > 0$ be small enough and $y = (x, \lambda) \in \Omega_{\varepsilon_0}$ such that $\|y - y^*\| = \varepsilon \le \varepsilon_0$.

Due to formulas (4.4) for the scaling parameter update and (4.18) from Lemma 3 we have

$$\frac{1}{\sqrt{L_2}}\varepsilon^{-\frac{1}{2}} \le k \le \frac{1}{\sqrt{L_1}}\varepsilon^{-\frac{1}{2}}. \tag{4.30}$$

We consider separately the active and the inactive constraints sets: I^* and I^0. We can rewrite system (4.10) as follows

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot)+\frac{1}{k^2}I_n & -\nabla c^T_{(r)}(\cdot) & -\nabla c^T_{(m-r)}(\cdot) \\ -k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\,\nabla c_{(r)}(\cdot) & I_r & 0 \\ -k\Lambda_{(m-r)}\Psi''_{(m-r)}(\cdot)\,\nabla c_{(m-r)}(\cdot) & 0 & I_{m-r} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \\ \Delta\lambda_{(m-r)} \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \bar\lambda_{(r)}-\lambda_{(r)} \\ \bar\lambda_{(m-r)}-\lambda_{(m-r)} \end{bmatrix}, \qquad (4.31)$$

where the second and the third systems of equations correspond to the active and the inactive sets respectively. First we consider the third system separately. After rearranging the terms we obtain

$$\hat\lambda_{(m-r)} := \lambda_{(m-r)} + \Delta\lambda_{(m-r)} = \bar\lambda_{(m-r)} + k\Lambda_{(m-r)}\Psi''_{(m-r)}(\cdot)\,\nabla c_{(m-r)}(\cdot)\Delta x.$$

Therefore for any $i \in I^0$ we have

$$\hat\lambda_i = \lambda_i + \Delta\lambda_i = \psi'(kc_i(x))\lambda_i + k\psi''(kc_i(x))\lambda_i\nabla c_i(x)^T\Delta x.$$

We remind that $\psi'(t) \le b_1 t^{-1}$, $-\psi''(t) \le b_2 t^{-2}$, $t \ge 0$, $b_1 \ge 0$, $b_2 \ge 0$. Also due to the standard second order optimality conditions and the boundedness Ω_{ε_0} there exists $\eta_1 > 0$, $\eta_2 > 0$, $\eta_3 > 0$ such that $c_i(x) \ge \eta_1$, $\|\nabla c_i(x)\| \le \eta_2$, $\|\Delta x\| \le \eta_3$, $i \in I^0$ if $(x,\lambda) \in \Omega_{\varepsilon_0}$. Using formula (4.4) for the scaling parameters update, keeping in mind that $|\lambda_i| \le \varepsilon$ for $i \in I^0$ and formula (4.30) we get

$$|\hat\lambda_i| \le \frac{b_1}{k\eta_1}\lambda_i + \frac{b_2\eta_2\eta_3}{k\eta_1^2}\lambda_i \le C_4\varepsilon^{\frac{3}{2}}, \quad i \in I^0, \qquad (4.32)$$

where $C_4 = \frac{b_1}{\sqrt{L_1}\eta_1} + \frac{b_2\eta_2\eta_3}{\sqrt{L_1}\eta_1^2}$.

Now we concentrate on the analysis of the primal-dual system, which corresponds to the active constraints. The first and the second equations of system (4.31) are equivalent to

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot)+\frac{1}{k^2}I_n & -\nabla c^T_{(r)}(\cdot) \\ -k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\,\nabla c_{(r)}(\cdot) & I_r \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) + \nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)} \\ \bar\lambda_{(r)}-\lambda_{(r)} \end{bmatrix}.$$

By multiplying the second equation of the system by $[-k\Lambda\Psi''(\cdot)]^{-1}$ we obtain

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot)+\frac{1}{k^2}I_n & -\nabla c^T_{(r)}(\cdot) \\ \nabla c_{(r)}(\cdot) & \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot)+\nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)} \\ \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1}(\bar{\lambda}_{(r)}-\lambda_{(r)}) \end{bmatrix}. \tag{4.33}$$

Keeping in mind that $c_i(x^*)=0$ for $i\in I^*$ and using the Lagrange formula we have

$$(\bar{\lambda}_i-\lambda_i)(-k\lambda_i\psi''(\cdot))^{-1}=(\lambda_i\psi'(kc_i(\cdot))-\lambda_i\psi'(kc_i(x^*)))\,(-k\lambda_i\psi''(\cdot))^{-1}=$$

$$\lambda_i k\psi''(\xi_i)(c_i(\cdot)-c_i(x^*))(-k\lambda_i\psi''(\cdot))^{-1}=-\psi''(\xi_i)(\psi''(\cdot))^{-1}c_i(\cdot),$$

where $\xi_i=k\theta_i c_i(\cdot)+k(1-\theta_i)c_i(x^*)=k\theta_i c_i(\cdot)$, $0<\theta_i<1$. Therefore the system (4.33) is equivalent to

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot)+\frac{1}{k^2}I_n & -\nabla c^T_{(r)}(\cdot) \\ \nabla c_{(r)}(\cdot) & \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot)+\nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)} \\ -\Psi''_{(r)}(\xi)\left[\Psi''_{(r)}(\cdot)\right]^{-1}c_{(r)}(\cdot) \end{bmatrix},$$

where $\Psi''_{(r)}(\xi)=diag(\psi''(\xi_i))_{i=1}^r$, or

$$B(\cdot)\Delta y_b=b(\cdot),$$

where

$$B(\cdot)=\begin{bmatrix} \nabla^2_{xx}L(\cdot)+\frac{1}{k^2}I_n & -\nabla c^T_{(r)}(\cdot) \\ \nabla c_{(r)}(\cdot) & \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1} \end{bmatrix},\quad b(\cdot)=\begin{bmatrix} -\nabla_x L(\cdot)+\nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)} \\ -\Psi''_{(r)}(\xi)\left[\Psi''_{(r)}(\cdot)\right]^{-1}c_{(r)}(\cdot) \end{bmatrix}$$

and $\Delta y_b=(\Delta x,\Delta\lambda_{(r)})$.

We are going to show that sequence generated by (4.11)–(4.12) is close to the one that generated bi Newton's method for the Lagrange system of equations that corresponds to the active constraints

$$\nabla L_{(r)}(x,\lambda_{(r)})=\nabla f(x)-\nabla c^T_{(r)}(x)\lambda_{(r)}=0, \tag{4.34}$$

$$c_{(r)}(x)=0. \tag{4.35}$$

By linearizing the equations (4.34)–(4.35) we obtain the following linear system for finding the Newton direction

$$\begin{bmatrix} \nabla^2_{xx}L_{(r)}(\cdot) & -\nabla c^T_{(r)}(\cdot) \\ \nabla c_{(r)}(\cdot) & 0 \end{bmatrix} \begin{bmatrix} \Delta x' \\ \Delta\lambda'_{(r)} \end{bmatrix} = \begin{bmatrix} -\nabla_x L_{(r)}(\cdot) \\ -c_{(r)}(\cdot) \end{bmatrix},$$

or

$$A(\cdot)\Delta y'_a=a(\cdot),$$

where

$$A(\cdot) = \begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T_{(r)}(\cdot) \\ \nabla c_{(r)}(\cdot) & 0 \end{bmatrix}, \quad a(\cdot) = \begin{bmatrix} -\nabla_x L_{(r)}(\cdot) \\ -c_{(r)}(\cdot) \end{bmatrix}$$

and $\Delta y'_a = (\Delta x', \Delta\lambda'_{(r)})$. The new primal-dual approximation is obtained by formulas

$$\hat{x}' = x + \Delta x', \quad \hat{\lambda}'_{(r)} = \lambda_{(r)} + \Delta\lambda'_{(r)}, \tag{4.36}$$

or

$$\hat{y}' = y + \Delta y'_a.$$

Let us estimate $\|\hat{y}_{(r)} - y^*_{(r)}\|$, where $\hat{y}_{(r)} = (\hat{x}, \hat{\lambda}_{(r)})$ is generated by (4.11)–(4.13).

$$\hat{y}_{(r)} - y^*_{(r)} = y_{(r)} + \Delta y_b - y^*_{(r)} = y_{(r)} + \Delta y'_a + \Delta y_b - \Delta y'_a - y^*_{(r)} = \hat{y}'_{(r)} - y^*_{(r)} - \Delta y'_a + \Delta y_b.$$

Therefore

$$\|\hat{y}_{(r)} - y^*_{(r)}\| \le \|y'_{(r)} - y^*_{(r)}\| + \|\Delta y'_a - \Delta y_b\| \tag{4.37}$$

First let us estimate $\|\Delta y'_a - \Delta y_b\|$. Due to Lemma 3 there exist inverse matrices $A^{-1} = A^{-1}(\cdot)$ and $B^{-1} = B^{-1}(\cdot)$ and for $a = a(\cdot)$, $b = b(\cdot)$ we have

$$\|\Delta y'_a - \Delta y_b\| = \|A^{-1}a - B^{-1}b\| = \|A^{-1}a - B^{-1}a + B^{-1}a - B^{-1}b\| =$$

$$\|\left(A^{-1} - B^{-1}\right)a + B^{-1}(a-b)\| \le \|A^{-1} - B^{-1}\|\|a\| + \|B^{-1}\|\|a - b\| \le$$

$$\|A^{-1}\|\|A - B\|\|B^{-1}\|\|a\| + \|B^{-1}\|\|a - b\|. \tag{4.38}$$

We consider the following matrix

$$A - B = \begin{bmatrix} \sum_{i=r+1}^m \lambda_i \nabla^2 c_i(x) - \frac{1}{k^2} I_n & 0 \\ 0 & -\frac{1}{k}\left[\Lambda_{(r)} \Psi''_{(r)}(\cdot)\right]^{-1} \end{bmatrix}$$

Due to formulas (4.4), (4.25) and (4.30) we obtain

$$|kc_i(\cdot)| \le \frac{C_3 L_2}{\sqrt{L_1}} \varepsilon^{\frac{1}{2}}, \quad i \in I^*. \tag{4.39}$$

and hence there is $\eta_4 > 0$ such that

$$|\lambda_i \psi''(kc_i(\cdot))| \ge \frac{1}{\eta_4}. \tag{4.40}$$

Due to boundedness of Ω_{ε_0} there exists $\tau_4 > 0$ such that for $y \in \Omega_{\varepsilon_0}$ we have

$$\|\nabla^2 c_i(x)\| < \tau_4, \quad i \in I_0. \tag{4.41}$$

Therefore keeping in mind formulas (4.4), (4.30) and (4.41) we have

$$\|A - B\| \le \max\left\{(\tau_4(m-r)+1)\varepsilon, \sqrt{L_2}\eta_4\varepsilon^{\frac{1}{2}}\right\} = \sqrt{L_2}\eta_4\varepsilon^{\frac{1}{2}} \tag{4.42}$$

for $0 < \varepsilon \le \varepsilon_0$ small enough.

Let us now estimate

$$\|a - b\| = \left\| \begin{matrix} -\nabla_x L_{(r)}(\cdot) + \nabla_x L(\cdot) - \nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)} \\ -c_{(r)}(\cdot) + (\Psi''(\xi))\left[\Psi''(kc_{(r)}(\cdot)\right]^{-1} c_{(r)}(\cdot) \end{matrix} \right\|. \tag{4.43}$$

For the first component we have

$$\| - \nabla_x L_{(r)}(\cdot) + \nabla_x L(\cdot) - \nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)}\| =$$

$$\| - \nabla_x L_{(r)}(\cdot) + \nabla_x L_{(r)}(\cdot) - \nabla c^T_{(m-r)}(\cdot)\lambda_{(m-r)} - \nabla c^T_{(m-r)}(\cdot)\Delta\lambda_{(m-r)}\| =$$

$$\|\nabla c^T_{(m-r)}(\cdot)(\lambda_{(m-r)} + \Delta\lambda_{(m-r)})\| = \|\nabla c^T_{(m-r)}(\cdot)\hat{\lambda}_{(m-r)}\| \le \eta_2 C_4 \varepsilon^{\frac{3}{2}}$$

For the second component of (4.43), using the Lagrange formula [1] for $i \in I^*$ we obtain

$$\left|\left(\frac{\psi''(\xi_i)}{\psi''(kc_i(\cdot))} - 1\right) c_i(\cdot)\right| \le \left|\frac{\psi''(\xi_i) - \psi''(kc_i(\cdot))}{\psi''(kc_i(\cdot))}\right| |c_i(\cdot)| \le \frac{|\psi'''(\bar{\xi}_i)||\xi_i - kc_i(\cdot)|}{|\psi''(kc_i(\cdot))|}|c_i(\cdot)| \le$$

$$\frac{|\psi'''(\bar{\xi}_i)||kc_i(\cdot)(\theta_i - 1)|}{|\psi''(kc_i(\cdot))|}|c_i(\cdot)|,$$

where $\bar{\xi}_i = \bar{\theta}_i\xi_i + k(1-\bar{\theta}_i)c_i(\cdot) = kc_i(\cdot)(\bar{\theta}_i\theta_i + 1 - \bar{\theta}_i)$. Due to (4.39) there exist $\eta_5 > 0$ such that for $i \in I^*$

$$|\psi'''(\bar{\xi}_i)| \le \eta_5.$$

Thus taking into consideration formulas (4.4), (4.24), (4.30), (4.39) and (4.40) we obtain for $i \in I^*$

$$\frac{|\psi'''(\bar{\xi}_i)||kc_i(\cdot)(1 - \theta_i)|}{|\psi''(kc_i(\cdot))|}|c_i(\cdot)| \le \eta_4\eta_5(1-\theta)C_3^2 L_2^2 L_1^{-\frac{1}{2}}\varepsilon^{\frac{3}{2}} = C_5\varepsilon^{\frac{3}{2}},$$

where $\theta = \min_{1\le i\le r}\theta_i$.

[1] Due to (4.39) we obtain $kc_i(x) > -0.5$ for $i \in I^*$. Therefore all the transformations described in section 3 infinitely times differentiable in Ω_{ε_0}.

Finally combining formulas (4.4), (4.19), (4.24), (4.25), (4.30), (4.38) and (4.42) we have

$$\|\Delta y'_a - \Delta y_b\| \le \|A^{-1}\|\|A-B\|\|B^{-1}\|\|a\| + \|B^{-1}\|\|a-b\| \le$$

$$4M^2\sqrt{L_2}\eta_4 \max\{C_2, C_3\} L_2 \varepsilon^{\frac{3}{2}} + 2M \max\{\eta_2 C_4, C_5\}\varepsilon^{\frac{3}{2}} = C_6 \varepsilon^{\frac{3}{2}}. \tag{4.44}$$

Due to quadratic convergence of Newton's method for solving Lagrange system of equations that corresponds to the active constraints (see [23]) from (4.36) we obtain

$$\|\hat{y}'_{(r)} - y^*_{(r)}\| \le C_0 \varepsilon^2, \tag{4.45}$$

where $\hat{y}'_{(r)} = (\hat{x}', \hat{\lambda}'_{(r)})$ and $y^*_{(r)} = (x^*, \lambda^*_{(r)})$.

Therefore combining (4.37), (4.44) and (4.45) we obtain

$$\|\hat{y}_{(r)} - y^*_{(r)}\| \le \|\hat{y}'_{(r)} - y^*_{(r)}\| + \|\Delta y'_a - \Delta y_b\| \le C_0\varepsilon^2 + C_6 \varepsilon^{\frac{3}{2}} \le C_7 \varepsilon^{\frac{3}{2}}. \tag{4.46}$$

Finally combining (4.32) and (4.46) for $\hat{y} = (\hat{x}, \hat{\lambda})$ we have

$$\|\hat{y} - y^*\| \le \max\{C_4, C_7\}\varepsilon^{\frac{3}{2}} = C\varepsilon^{\frac{3}{2}} = C\|y - y^*\|^{\frac{3}{2}}.$$

The proof of Theorem 1 is complete.

Remark 1. It is well known (see [23]) that under the standard second order optimality conditions Newton's method for Lagrange system of equations associated with equality constraints locally generates primal-dual system, that converges to the primal-dual solution quadratically. This fact played an important role in our convergence proof. A number of penalty and augmented Lagrangian type methods that asymptotically produce primal-dual direction close to the Newton direction were considered in [8,9,12–14,34], where asymptotic linear and superlinear rates of convergence were observed under the standard second order optimality conditions.

Remark 2. It follows from the proof that the regularization term $\frac{1}{k^2} I_n$ does not compromise the rate of convergence. On the other hand, we will see in the next section that this term also insures the global convergence of the algorithm.

Remark 3. The proof of Theorem 1does not require convexity assumptions. The convexity of function $f(x)$ and concavity of function $c_i(x)$ are used in the next section to prove the global convergence of the PDNRD method.

5. Globally convergent PDNRD method

In this section we describe a globally convergent primal-dual NR algorithm with dynamic scaling parameter update that has the asymptotic 1.5-Q-superlinear rate of convergence. At each iteration the algorithm solves system (4.11). If the primal-dual direction $(\Delta x, \Delta\lambda)$ does not produce the superlinear reduction of the merit function, then the primal direction Δx is used for minimization $\mathcal{L}(x, \lambda, k)$ in x. Therefore convergence to the neighborhood of the primal-dual solution is guaranteed by NR method (3.9)–(3.10). Eventually in the neighborhood of the primal-dual solution due to Theorem 1 one step of the PDNRD method (4.11) –(4.13) will be enough to obtain the desired 1.5-Q-superlinear reduction of the merit function and the bound (4.29) will take place. Figures 1 and 2 describe the PDNRD algorithm. It has some similarities to the globally convergent Newton's method for unconstrained optimization when the Newton direction with steplength is used at each step to guarantee convergence. From some point on the steplength becomes equal one and "pure" Newton's method converges with quadratic rate.

Step 1: Initialization:

An initial primal approximation $x^0 \in \mathbb{R}^n$ is given.

An accuracy parameter $\varepsilon > 0$ and the initial scaling paramter $k > 0$ are given.

Parameters $\alpha > 1$, $0 < \gamma < 1$, $0 < \eta < 0.5$ $\sigma > 0$, $\theta > 0$ are given.

Set $x := x^0$, $\lambda^0 := (1, \dots, 1) \in \mathbb{R}^m$, $r := \nu(x, \lambda)$, $\lambda_g := \lambda^0$.

Step 2: If $r \le \varepsilon$, Stop, **Output**: x, λ.

Step 3: Find direction: $(\Delta x, \Delta\lambda) := \mathit{PrimalDualDirection}(x, \lambda)$.

Set $\hat{x} := x + \Delta x$, $\hat{\lambda} := \lambda + \Delta\lambda$.

Step 4: If $\nu(\hat{x}, \hat{\lambda}) \le \min\{r^{\frac{3}{2}-\theta}, 1-\theta\}$, Set $x := \hat{x}$, $\lambda := \hat{\lambda}$, $r := \nu(x, \lambda)$, $k := \max\{\frac{1}{\sqrt{r}}, k\}$, Goto Step 2.

Step 5: Decrease $t \le 1$ until $\mathcal{L}(x + t\Delta x, \lambda_g, k) - \mathcal{L}(x, \lambda_g, k) \le \eta t (\nabla\mathcal{L}(x, \lambda_g, k), \Delta x)$

Step 6: Set $x := x + t\Delta x$, $\hat{\lambda} := \lambda_g \psi'(kc(x + t\Delta x))$.

Step 7: If $\|\nabla_x \mathcal{L}(x, \lambda_g, k)\| \le \frac{\sigma}{k}\|\hat{\lambda} - \lambda_g\|$, Goto Step 9.

Step 8: Find direction: $(\Delta x, \Delta\lambda) := \mathit{PrimalDualDirection}(x, \lambda_g)$, Goto Step 5.

Step 9: If $\nu(x, \hat{\lambda}) \le \gamma r$, Set $\lambda := \hat{\lambda}$, $\lambda_g := \hat{\lambda}$, $r := \nu(x, \lambda)$, $k := \max\{\frac{1}{\sqrt{r}}, k\}$, Goto Step 2.

Step 10: Set $k := k\alpha$, Goto Step 8.

Fig. 1. PDNRD algorithm.

function $(\Delta x, \Delta\lambda) := PrimalDualDirection(x, \lambda)$

begin

$\bar{\lambda} := \psi'(kc(x))\lambda$

Solve PDNR system:

$$\begin{bmatrix} \nabla^2_{xx} L(x,\lambda) + \frac{1}{k^2} I & -\nabla c^T(x) \\ -k\Psi''(kc(x))\,\Lambda\nabla c(x) & I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x,\lambda) \\ \bar{\lambda} - \lambda \end{bmatrix},$$

where $\Psi''(kc(x)) = diag\ (\psi''(kc_i(x)))_{i=1}^m$, $\Lambda = diag\ (\lambda_i)_{i=1}^m$

end

Fig. 2. Newton PDNRD direction.

The following lemma guarantees global convergence of the PDNRD method. Let us consider the iterative method

$$N_k(x^s, \lambda) \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x^s, \lambda) \\ \bar{\lambda} - \lambda \end{bmatrix}, \tag{5.1}$$

$$\alpha_s : \quad \mathcal{L}(x + \alpha_s \Delta x, \lambda, k) - \mathcal{L}(x, \lambda, k) \le \eta\alpha_s(\nabla\mathcal{L}(x, \lambda, k), \Delta x), \tag{5.2}$$

where $0 < \eta < 1$.

$$x^{s+1} := x^s + \alpha_s \Delta x^s. \tag{5.3}$$

We find the steplength α_s from P. Wolfe's condition [22].

Lemma 4. *For any primal-dual pair $(x, \lambda) \notin \Omega_{\varepsilon_0}$ and fixed $\lambda \in \mathbb{R}^m_{++}$ and $k > 0$ method (5.1)–(5.3) generates the primal-dual sequence that converges to the unconstrained minimizer of $\mathcal{L}(x, \lambda, k)$ in x.*

Proof. We can rewrite system (4.11) as follows.

$$\left(\nabla^2_{xx} L(x^s, \lambda) + \frac{1}{k^2} I_n)\right) \Delta x^s - \nabla c(x^s)^T \Delta\lambda = -\nabla_x L(x^s, \lambda) = -\nabla_x L(\cdot), \tag{5.4}$$

$$-k\Lambda\Psi''(kc(x^s))\,\nabla c(x^s)\Delta x^s + \Delta\lambda = \bar{\lambda} - \lambda, \tag{5.5}$$

where $\bar{\lambda} = \lambda\Psi'(kc(x^s))$.

By substituting the value of $\Delta\lambda$ from (5.5) into (5.4) the primal Newton direction one can find from the following system

$$M(x^s,\lambda,k)\Delta x^s = -\nabla_x L(x^s,\bar{\lambda}) = -\nabla_x \mathcal{L}(x^s,\lambda,k),$$

where

$$M(\cdot) = M(x^s,\lambda,k) = \nabla^2_{xx}L(x^s,\lambda) + \frac{1}{k^2}I_n - k\nabla c^T(x^s)\Psi''\left(kc(x^s)\right)\Lambda\nabla c(x^s).$$

Due to formula (4.4) the scaling parameter $k>0$ increases unboundedly only when (x,λ) approaches the solution. It means there is a bound $\bar{k}>0$ such that for any $(x,\lambda)\notin\Omega_{\varepsilon_0}$ the scaling parameter $k\le\bar{k}$.

Due to the convexity of $f(x)$, concavity of $c_i(x)$,, transformation properties 4^0, and the regularization term $\frac{1}{k^2}I_n$ the matrix $M(\cdot)$ is positive definite with uniformly bounded condition number for all s, and $k\le\bar{k}$,, i.e.

$$m_1(x,x)\le(M(\cdot)x,x)\le m_2(x,x)\quad,\forall x\in\mathbb{R}^n$$

and $0<m_1<m_2$. It implies that Δx^s is a descent direction for minimization of $\mathcal{L}(x,\lambda,k)$ in x.. In addition if the steplength satisfies Wolfe's condition we have $\lim_{s\to\infty}\|\nabla_x\mathcal{L}(x^s,\lambda,k)\|=0$ (see e.g.[22]). Therefore for any given $(\lambda,k)\in\mathbb{R}^{m+1}_{++}$ the sequence generated by method (5.1)–(5.3) converges to the unconstrained minimizer of $\mathcal{L}(x,\lambda,k)$ in x. Lemma 4 is proven.

Combining the results of Proposition 1, Theorem 1 and Lemma 4 together we obtain the following.

Theorem 2. *Under the assumptions of Theorem 1 the PDNRD method generates the primal-dual sequence that converges to the primal-dual solution with asymptotic 1.5-Q-superlinear rate of convergence.*

The PDNRD algorithm uses the primal Newton direction with steplength and following the NR path until the primal-dual approximation reaches the neighborhood Ω_{ε_0}. Then due to Theorem 1 it requires at most $\mathcal{O}(\log\log\varepsilon^{-1})$ steps to find the primal-dual solution with accuracy $\varepsilon>0$. The neighborhood Ω_{ε_0} is unknown a priory. Therefore the PDNRD may switch to Newton's method for primal-dual system (4.5)-(4.6) prematurely, before (x,λ) reaches Ω_{ε_0}. Then PDNRD algorithm will recognize this in less than $\mathcal{O}(\log\log\varepsilon^{-1})$ steps and will continue following the NR trajectory with much bigger value of scaling parameter $k>0$.

The algorithm has been implemented in MATLAB and tested on a variety of problems. The following section presents some numerical results from the testing.

6. Numerical results

Tables 1–12 present the performance of the PDNRD method on some problems from COPS [5] and CUTE [6] sets. In our calculations we use transformation ψ_{q_2} introduced in section 3. We show the number of variables n and constraints m. Then we show the objective function value, the norm of the gradient of the Lagrangian, the complementarity violation, the primal-dual infeasibility, and the number of Newton steps required to reduce the merit function by an order of magnitude.

One of the most important observations following from the obtained numerical results is significant acceleration of convergence of the PDNRD method when the primal-dual sequence approaches the solution. It is in full correspondence with Theorem 1. For all problems we observed the "hot" start phenomenon: from some point on only one Newton step is required to reduce the value of the merit function at least by the order of magnitude.

Table 1. CUTE, aircrfta: $n = 5$, $m = 5$, linear objective, quadratic constraints

it	f	$\|\nabla L(x, \lambda)\|$	gap	constr violat	# of steps
0	4.031e+02	6.5281e+03	0.0000e+00	2.8370e+00	0
1	2.709e-07	6.1602e-02	1.8062e-07	3.7308e-05	4
2	3.974e-08	1.3987e-06	7.6025e-14	1.4923e-05	1
3	9.523e-16	1.7577e-10	2.6191e-20	9.2782e-11	1
Total number of Newton steps					6

Table 2. CUTE, airport: $n = 84$, $m = 210$, linear objective, quadratic constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	1.345e+07	1.6597e+06	3.4666e+02	1.0360e+02	0
1	4.801e+04	1.0268e+03	7.2841e+01	2.9730e-03	12
2	4.795e+04	2.2093e+00	1.3615e+00	3.0112e-07	13
3	4.795e+04	1.0899e-03	4.9758e-02	7.0738e-08	3
4	4.795e+04	3.9039e-05	1.0174e-03	2.6057e-09	2
5	4.795e+04	3.2575e-08	2.9585e-10	3.9753e-11	1
6	4.795e+04	7.1054e-12	5.5471e-14	1.7153e-14	1
Total number of Newton steps					32

Table 3. CUTE, avgasa: $n = 6$, $m = 18$, quadratic objective, linear constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	-4.071e+00	4.4312e+00	9.0000e+00	-0.0000e+00	0
1	-4.175e+00	1.8651e-03	1.5612e-01	3.9168e-03	17
2	-4.169e+00	5.0646e-03	3.7190e-03	2.8939e-04	10
1	-4.169e+00	1.5713e-09	3.4780e-05	3.5657e-05	2
2	-4.169e+00	1.4853e-08	3.0974e-06	1.3140e-05	1
3	-4.169e+00	2.5863e-09	3.4712e-07	3.9769e-06	1
4	-4.169e+00	1.4468e-10	2.5983e-08	4.7697e-07	1
5	-4.169e+00	1.7776e-12	7.9031e-10	1.8035e-08	1
6	-4.169e+00	2.5008e-15	5.1032e-12	1.2809e-10	1
7	-4.169e+00	4.2327e-16	2.9326e-15	7.5273e-14	1
Total number of Newton steps					35

Table 4. COPS: Journal bearing: $n = 5000$, $m = 5000$, nonlinear objective, bounds

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	-4.504e+02	5.1364e+01	1.6229e+03	0.0000e+00	0
1	-8.002e-02	9.1922e-08	9.2602e-03	7.0564e-03	20
2	-1.550e-01	9.3995e-13	1.6896e-05	3.4093e-05	7
3	-1.550e-01	1.4592e-15	9.1966e-09	6.0043e-07	4
4	-1.550e-01	7.7398e-17	1.1702e-10	1.3002e-08	2
5	-1.550e-01	6.2450e-17	1.5082e-11	2.7984e-09	1
6	-1.550e-01	6.5919e-17	1.1229e-12	4.4897e-10	1
7	-1.550e-01	6.5919e-17	6.5135e-14	5.9776e-11	1
8	-1.550e-01	6.2450e-17	3.9621e-15	6.6580e-12	1
Total number of Newton steps					37

Table 5. CUTE: biggsb2: $n = 1000$, $m = 1998$, quadratic objective, bounds

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	-4.797e+00	3.1761e+01	8.9910e+02	-0.0000e+00	0
1	2.566e-02	9.8864e-02	1.7369e-05	7.6468e-04	12
2	1.500e-02	1.5879e-13	2.0104e-06	1.2438e-05	4
3	1.500e-02	3.3544e-14	3.1460e-07	2.6468e-06	1
4	1.500e-02	3.9262e-15	1.8169e-08	1.8039e-07	1
5	1.500e-02	4.4409e-16	1.2791e-10	1.3011e-09	1
6	1.500e-02	4.1517e-16	3.1744e-14	4.2613e-12	1
Total number of Newton steps					20

Table 6. CUTE: congigmz: $n = 3$, $m = 5$, minimax: linear objective, nonlinear constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	1.112e+04	8.5085e+03	3.6000e+01	1.4000e+01	0
1	3.321e+01	1.5855e+00	6.8324e+00	0.0000e+00	20
2	3.412e+01	6.5803e-03	7.5631e-01	1.4072e-01	7
3	2.800e+01	1.2877e-02	1.0537e-03	1.2996e-04	4
4	2.800e+01	3.9900e-07	2.8043e-05	1.5627e-06	2
5	2.800e+01	7.8753e-10	1.3920e-09	4.7149e-08	1
6	2.800e+01	1.9984e-15	3.9972e-13	1.0409e-11	1
7	2.800e+01	1.7764e-15	8.2580e-16	0.0000e+00	1
Total number of Newton steps					36

Table 7. CUTE: dtoc5: $n = 998$, $m = 499$, minimax: quadratic objective, quadratic constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	5.020e+02	1.0020e+03	0.0000e+00	1.0020e+00	0
1	1.536e+00	6.5637e-07	2.3048e-03	1.1068e-04	5
2	1.535e+00	2.1110e-09	3.3520e-06	1.6171e-07	2
3	1.535e+00	4.7321e-13	3.3241e-07	1.6556e-09	1
4	1.535e+00	4.6872e-15	2.0525e-08	1.0679e-10	1
5	1.535e+00	1.1015e-15	8.4896e-10	4.7902e-12	1
6	1.535e+00	1.7208e-15	1.7808e-11	1.0170e-13	1
7	1.535e+00	1.2546e-15	5.4923e-14	4.2414e-16	1
Total number of Newton steps					12

Table 8. CUTE: gilbert: $n = 1000$, $m = 1$, quadratic objective, quadratic constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	1.250e+10	1.5811e+08	0.0000e+00	5.0000e+04	0
1	4.821e+02	1.3971e-07	1.5065e+00	9.3099e-02	25
2	4.820e+02	9.2648e-07	1.3036e-01	7.4314e-03	5
3	4.820e+02	2.1307e-04	2.5920e-02	1.4683e-03	1
4	4.820e+02	1.5072e-05	3.0774e-03	1.7412e-04	1
5	4.820e+02	3.4087e-07	2.7104e-04	1.5334e-05	1
6	4.820e+02	1.8024e-10	3.6471e-08	9.9661e-07	1
7	4.820e+02	8.6531e-13	6.5325e-10	1.7851e-08	1
8	4.820e+02	5.5511e-16	1.5904e-12	4.3460e-11	1
9	4.820e+02	3.3307e-16	1.7876e-16	4.8850e-15	1
Total number of Newton steps					37

Table 9. Hock & Schittkowski 117: $n = 15$, $m = 20$, quadratic objective, quadratic constraints and bounds

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	5.436e+04	4.9754e+04	1.0800e+02	3.6000e+01	0
1	4.212e+02	7.6243e+00	9.6041e+00	5.5393e+00	62
2	3.235e+01	1.3047e-03	3.4444e-03	6.6456e-05	27
3	3.235e+01	3.2383e-07	1.5505e-04	8.0089e-06	2
4	3.235e+01	2.9211e-08	1.7757e-07	6.4741e-07	1
5	3.235e+01	2.2220e-11	8.2702e-10	5.5118e-09	1
6	3.235e+01	7.1054e-15	4.9593e-13	3.9804e-12	1
Total number of Newton steps					94

Table 10. CUTE: optctrl6 $n = 118$, $m = 80$, quadratic objective, quadratic constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	1.610e+06	2.4981e+06	1.5601e+06	9.8000e+00	0
1	1.937e+03	2.3443e-02	5.1348e+02	1.3507e-02	13
2	2.048e+03	9.4961e-09	8.8005e+01	1.5054e-03	12
3	2.048e+03	3.2521e-06	2.7358e+00	4.4311e-05	7
4	2.048e+03	2.4424e-08	4.6971e-01	7.5977e-06	2
5	2.048e+03	3.3227e-09	5.7014e-02	9.2203e-07	2
6	2.048e+03	3.6343e-09	4.8080e-03	7.7753e-08	2
7	2.048e+03	2.6724e-07	2.7814e-04	4.4978e-09	1
8	2.048e+03	2.5466e-11	5.3345e-09	1.7679e-10	1
9	2.048e+03	2.5580e-11	2.0977e-10	6.9508e-12	1
10	2.048e+03	2.7057e-11	8.2480e-12	2.7331e-13	1
Total number of Newton steps					42

Table 11. CUTE: optmass $n = 126$, $m = 105$, quadratic objective, linear and quadratic constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	-9.642e-01	1.0025e+00	2.1000e+01	1.0000e-02	0
1	3.233e-02	2.8648e-06	2.7441e-01	8.1574e-03	91
2	-1.511e-01	1.3159e-04	6.6208e-03	9.0302e-04	25
3	-1.517e-01	2.8382e-05	4.6813e-04	5.9219e-05	7
4	-1.517e-01	7.2289e-09	1.0644e-05	2.7303e-05	1
5	-1.517e-01	3.0537e-11	3.5932e-07	1.5766e-06	1
6	-1.517e-01	1.1965e-12	7.8788e-09	8.6149e-09	1
7	-1.517e-01	2.0487e-15	5.0968e-11	2.6272e-12	1
8	-1.517e-01	5.5511e-17	2.5146e-14	1.3461e-15	1
Total number of Newton steps					128

Table 12. COPS: Isometrization of α -pinene $n = 4000$, $m = 4000$, nonlinear objective, nonlinear constraints

it	f	$\|\nabla L(x,\lambda)\|$	gap	constr violat	# of steps
0	1.096e+10	9.6378e+10	0.0000e+00	2.3600e+01	0
1	2.095e+01	6.1850e-04	2.6426e+00	4.1064e-05	17
2	1.989e+01	1.0998e-01	2.5621e-01	2.8963e-06	4
3	1.987e+01	2.9277e+00	1.6708e-02	2.4864e-07	2
4	1.987e+01	3.0175e-04	9.4867e-04	2.1777e-08	2
1	1.987e+01	1.2649e-03	2.1393e-06	1.3047e-09	1
2	1.987e+01	4.4104e-06	1.1108e-07	7.1941e-11	1
3	1.987e+01	1.4076e-08	5.5255e-09	3.6255e-12	1
4	1.987e+01	1.5019e-09	2.5360e-10	1.6685e-13	1
Total number of Newton steps					29

7. Concluding remarks

Theoretical and numerical results obtained for the PDNRD method emphasize the fundamental difference between the primal-dual NR approach and Newton NR methods [19], [26], which are based on sequential unconstrained minimization $\mathcal{L}(x, \lambda, k)$ followed by the Lagrange multipliers update. The Newton NR method converges globally with a fixed scaling parameter, keeps stable the Newton area for the unconstrained minimization and allows the observation of the "hot start" phenomenon [19], [24]. It leads to asymptotic linear convergence with a given factor $0 < \gamma < 1$ in one Newton step. However, the unbounded increase of the scaling parameter compromises convergence, since the Newton area for unconstrained minimization shrinks to a point. Moreover, in the framework of the NR method, any drastic increase of the scaling parameter after the Lagrange multipliers update leads to a substantial increase of the computational work per update because several Newton steps are required to get back to the NR trajectory. The situation is fundamentally different with Newton's method for the primal-dual system (4.5)-(4.6) in the neighborhood of Ω_{ε_0}. The drastic increase of the scaling parameter does not increase the computational work per step. Just the opposite: by using (4.13) for the scaling parameter update we obtain the Newton direction for the primal-dual system (4.5)-(4.6) close to the Newton direction for the Lagrange system of equations that corresponds to the active set. The latter direction guarantees the quadratic convergence of the corresponding primal-dual sequences ([23]). Therefore the PDNRD uses the best properties of both Newton's NR method far from the solution and Newton's method for the primal-dual system (4.5)-(4.6) in the neighborhood of the solution. At the same time PDNRD is free from their fundamental drawbacks.

The PDNRD method recalls the situation in unconstrained smooth optimization, in which Newton's method with steplength is used to guarantee global convergence. Locally the steplength automatically becomes equal one and Newton's method gains the asymptotic quadratic convergence.

A few important issues remain for future research. The NR multipliers method with inverse proportional scaling parameter update [28] generates such a primal-dual sequence that the Lagrange multipliers corresponding to the inactive constraints converge quadratically to zero. This fact can be used to eliminate the inactive constraints in the early stage of the computational process. Then the PDNRD method evolves into Newton's method for the Lagrange system of equations that corresponds to the active constraints. Therefore under the standard second order optimality conditions, the PDNRD method has a potential to be augmented to a globally convergent method with asymptotic quadratic rate.

Another important issue is the generalization of the PDNRD method for nonconvex problems.

Also, more work should be done to find an efficient way of solving the PD system (4.11) that accounts for the system's special structure.

Acknowledgements. We thank the anonymous referees for their valuable remarks, which contributed to improvement of the paper.

References

1. A. Auslender, R. Comminetti, M. Haddou *Asymptotic analysis of penalty and barrier methods in convex and linear programming,* Mathematics of Operations Research **22**, (1997), 43–62.
2. M. Bendsoe, A. Ben-Tal, J. Zowe (1994): Optimization Methods for Truss Geometry and Topology Design, Structural Optimization **7**, 141–159
3. A. Ben-Tal, I. Yuzefovich and M. Zibulevski (1992): Penalty/Barrier Multiplier Method for minimax and constrained smooth convex problems, Research Report 9/92, Optimization Laboratory, Faculty of Industrial Engineering and Management, Technion
4. A. Ben-Tal, M. Zibulevski (1997): Penalty-barrier methods for convex programming problems, SIAM Journal on Optimization, **7**, 347–366
5. A.S. Bondarenko, D.M. Bortz, J.J. More (1999): COPS: Large-scale nonlinearly constrained optimization problems, Mathematics and Computer Science Division, Argonne National Laboratory, Technical Report ANL/MCS-TM-237
6. I. Bongartz, A.R. Conn, N. Gould, Ph.L. Toint, *Constrained and unconstrained testing environment*, www.cse.clrc.ac.uk/Activity/CUTE+74
7. M. Breitfelt, D. Shanno (1996): Experience with modified log-barrier method for nonlinear programming, Annals of Operations Research, **62**, 439–464
8. R. Cominetti and J.-P. Dussault (1994): A stable exponential-penalty algorithm with superlinear convergence, Journal of Optimization Theory and Applications, **83**
9. J.-P. Dussault (1998): Augmented penalty algorithms, I.M.A. Journal on Numerical Analysis, **18**, 355-372
10. A. Forsgren, P. Gill (1998): Primal-dual interior methods for nonconvex nonlinear programming, SIAM Journal on Optimization **8**, 1132–1152
11. D. M. Gay and M. L. Overton and M. H. Wright (1998): A primal-dual interior method for nonconvex nonlinear programming, A primal-dual interior method for nonconvex nonlinear programming, Advances in Nonlinear Programming, Ed. Y. Yuan, 31–56, Kluwer Academic Publishers, Dordrecht.

12. N.I.M. Gould (1986): On the accurate determination of search directions for simple differentiable penalty functions, I.M.A. Journal on Numerical Analysis, **6**, 357–372

13. N.I.M. Gould (1989): On the convergence of a sequential penalty function method for constrained minimization, SIAM Journal on Numerical Analysis, **26**, 107–108

14. N.I.M. Gould, D. Orban, A. Sartenaer, and P. Toint (2001): Superlinear convergence of primal-dual interior point algorithms for nonlinear programming, SIAM Journal on Optimization, **11**, 974–1002

15. B.W. Kort, D.P. Bertsekas (1973): Multiplier methods for convex programming, Proceedings, IEEE Conference on Decision and Control, San Diego, California, 428–432

16. I. Lustig, R. Marsten, D. Shanno (1994): Computational experience with globally convergent primal-dual predictor-corrector algorithm for linear programming, Mathematical Programming **66**, 23–135

17. N. Megiddo (1989): Pathways to the optimal set in linear programming, in N. Megiddo, ed., Interior Point and Related Methods, Springer-Verlag, New York , Ch. 8, 131–158

18. S. Mehrotra (1992): On implementation of a primal-dual interior point method, SIAM Journal on Optimization **2**, 575–601.

19. A. Melman, R. Polyak (1996), The Newton Modified Barrier Method for QP Problems, Annals of Operations Research, **62**, 465–519

20. S. Nash, R. Polyak and A. Sofer (1994): A numerical comparison of Barrier and Modified Barrier Method for Large-Scale Bound-Constrained Optimization, in "Large Scale Optimization, state of the art", Ed. by W. Hager, D. Hearn and P. Pardalos, Kluwer Academic Publisher , 319–338

21. E. Ng, B.W. Peyton (1993): Block sparse Cholesky algorithms on advanced uniprocessor computers, SIAM Journal of Scientific Computing, **14**, 1034–1056

22. J. Nocedal, S. Wright (1999): Numerical Optimization, Springer, New York

23. B.T. Polyak (1970): Iterative methods for optimization problems with equality constraints using Lagrange multipliers, Journal of computational mathematics and mathematical physics, **10**, 1098–1106

24. R. Polyak (1992): Modified Barrier Functions Theory and Methods, Mathematical Programming **54**, 177–222

25. R. Polyak, M. Teboulle (1997): Nonlinear Rescaling and Proximal-like Methods in convex optimization, Mathematical Programming **76**, 265–284

26. R. Polyak, I. Griva, J. Sobieski (1998): The Newton Log-Sigmoid method in Constrained Optimization, A Collection of Technical Papers, 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization **3**, 2193–2201

27. R. Polyak (2001): Log-Sigmoid Multipliers Method in Constrained Optimization, Annals of Operations Research **101**, 427–460

28. R. Polyak (2002): Nonlinear rescaling vs. smoothing technique in convex optimization, Mathematical Programming **92**, 197–235

29. R. Polyak, I. Griva, Primal-Dual Nonlinear Rescaling method for convex optimization, Techncal Report 2002-05-01, George Mason University, to appear in JOTA in July 2004, (http://mason.gmu.edu/ rpolyak/papers.html).

30. D. Shanno, R. Vanderbei (1999): An interior-point algorithm for nonconvex nonlinear programming, COAP **13**, 231–252

31. R. Vanderbei (1995): Symmetric Quasidefinite Matrices SIAM Journal on Optimization, **5**, 100–113

32. S. Wright (1997): Primal-Dual Interior Points methods, SIAM

33. Y. Zhang (1996): Solving Large-Scale Linear Programs by Interior-Point Methods Under MATLAB Environment, Dept. of Computational and Applied Mathmatics, Rice University, Houston

34. Y. Zhang, R. Tapia, J. Dennis (1992): On the superlinear and quadratic convergence at primal-dual interior point linear programming algorithms, SIAM Journal on Optimization **2**, 304–324